

Effectiveness of the Guide for Monitoring Child Development Home-Based Parenting Support Intervention for Early Childhood Development: A Cluster-Randomized Trial

Statistical Analysis Plan

NCT Number

NCT04665297

Secondary Registration

CTRI/2020/12/029748

Document Version

Version 5 (Final)

Document Date

February 18, 2026

Statistical Analysis Plan, Version 5

Hypothesis Testing

In our protocol, we have one primary hypothesis and two secondary hypotheses. Our main efficacy endpoint is to assess Bayley-4 Language composite scores at 12 months by study arm. For each of the main hypotheses, we will also test separately by site (India or Guatemala).

	Sample	12 month endpoint (Main efficacy)*
1	Primary Full sample	H_{01} : There is no difference in Bayley-4 language composite scores at 12 months by study arm
2	Secondary Full sample	H_{02} : There is no difference in Bayley-4 other composite scores (motor, cognitive, total) at 12 months by study arm
3	Secondary Full sample	H_{03} : There is no difference in total HOME scores at 12 months by study arm

Note: all composite Bayley-4 scores are by definition age-adjusted.

*All of these will also be done separately by site.

Analysis Assumptions

Unit of inference: individual. Answering the question, "What would the GMCD response be for the average person" (not cluster).

Unit of analysis: individual

Analysis type: Conditional. Specifically, conditional on cluster-level variables of X level of community stunting and Y distance from health center. Answering the question, "What would the GMCD response be for the average person with X level of community stunting and Y distance from health center?"

Analysis plan: mixed effects regression, controlling for baseline score in child (k), nested in frontline worker (j), nested in NGO or Anganwadi Center (i). Both Guatemala and India have instances when more than one frontline worker is within the NGO/AC. We will assess each level of nesting; if both are contributing to variance, we will include both. If that is not the case, we will only include the main cluster level. We expect to analyze data in Stata, but some analyses may be conducted in other programs, such as R.

We plan to condition only on the factors used to restrict randomization: the proportion of stunting in under 5 children per cluster, and the mean distance from health services per cluster. We will conduct both pooled analyses and stratified by site (India or Guatemala). As sensitivity analyses, we will assess effect modification for child sex, and age category and other factors (found in the possible adjustment table) by including an interaction term with the intervention.

Basis for analysis: intention-to-treat framework.

Individuals and clusters who were lost to follow-up will have as much data as exists analyzed. In the event of missing data, we will assess missingness and whether treatment arm or other primary predictors (specifically cluster-level variables and sex of child) are associated with missingness of data. We will handle missing data by analysis using Maximum Likelihood if only the outcome data are missing. In the event that primary predictor data are missing, we will assess the amount of missing data. If 10% or less, we will handle missing data through either Full Information Maximum Likelihood using Stata's SEM commands or multiple imputation prior to analysis.

Description of Variables

Outcome/dependent variables

Variable name	Description
bayley_lang	brawreclang + brawexplang
bayley_z	Age-adjusted z score based on study population
Bayleytotal	Total raw score for Bayley 4 (fine motor + gross motor + rec lang + exp lang + cognition)
bayley_motor	Total raw score fine motor + gross motor
Bayley_std_score_lan	Standardized scaled language score (rec_lang + exp_lang scaled scores)
Bayley_std_score_cog	Standardized scaled cognitive score
Bayley_std_score_mot	Standardized scaled motor score (fine motor + gross_motor)
Home_prop	Proportion of "yes" answers out of total possible "yes" answers for HOME-IT and HOME-EC
home_total	Total raw score for HOME-IT & HOME-EC
responsivity_2	Total responsivity raw score for HOME-IT
acceptance_2	Total acceptance raw score for HOME-IT
organization_2	Total organization raw score for HOME-IT
learning_materials_2	Total learning materials raw score for HOME-IT
involvement_2	Total involvement raw score for HOME-IT
variety_2	Total variety raw score for HOME-IT
learning_stimulation	Stimulation score for HOME-EC
environment	Physical environment for HOME-EC tool
academic1	Academic stimulation for HOME-EC tool
modeling2	Modeling for HOME-EC tool

Stratification/Cluster variables

Variable name	Description
site	India, Guatemala
cluster	Coded variable, for clinics in India and clinics/FLW in Guatemala

CHW	Coded variable, for Anganwadi workers in India and FLW in Guatemala
age_month_dichot	Baby's age in months, dichotomized at 0–12 and 13–<25 at enrollment
clinic	NGO code or Anganwadi center code

Independent variables

Variable name	Description
arm	Study arm: 1 = (masked interv or control), 0 = (masked control or intervention), 3 = withdrawn
study_arm	1 = unmasked interv or control, 0 = unmasked control or interv; using intention to treat baseline assignment. "Withdrawn" need to be included in either 1 or 2 (not yet created)

Possible control/adjustment variables

Variable name	Description	Time varying?	To be created?
Edadmeses	Baby age, in months, calculated from days, at enrollment	N	N
birth_wt_grams	Baby's birthweight, in grams, continuous	N	N
Lbw	Baby's birthweight, categorized into WHO categories of normal ≥ 2500 g, low < 2500 g, and very low < 1500 g	N	Y
Maternal age	Reported age at birth of enrolled child in years (current age in years minus months of child at enrollment, rounded to years)	N	Y
hb, hb2	Anemia at baseline	N	N
anemia_baseline	Baseline hemoglobin, < 10.5 vs. ≥ 10.5 (adjusted for Guatemala altitude) if child is under 24 months; < 11 vs. ≥ 11 if child is 24 months or older, based on 2024 WHO guidelines	N	Y
child_5y	Number of children under 5 in household at baseline, continuous	N	N
child_5y_cat	Number of children under 5 in household at baseline: 1, 2–3, ≥ 4	N	Y
school_y	Years of schooling of caregiver at baseline, continuous	N	N
school_y_cat	Years of schooling of caregiver at baseline, categorized into no education, primary, secondary or higher	N	Y
household wealth	Calculated from assets and poverty tool, baseline	N	Y
stunting	Baseline status, measured height or length by age. Dichotomized at ≤ -2 vs. > -2	N	Y
Haz06_b	Baseline height/length for age z score	N	Y
Sex	1 = male, 2 = female	N	N

Note: the definition of "primary school" is different in India and Guatemala. India primary school = 1–8, secondary = 9–12. Guatemala primary school = 1–6, lower secondary = 7–9, upper secondary = 10–12 or 10–11. We will categorize each country separately based on the years of "compulsory" primary education.

Statistical Methods

Data cleaning

Cleaning will be performed by the data managers at each site: at present that would be Scott for Guatemala, and Ashwini for India.

Analysis

Summary baseline statistics: We will conduct exploratory data analysis on cluster-level, household level and individual level factors. We will provide means and standard deviations (or medians and interquartile ranges) for continuous variables and frequencies and percents for dichotomous and categorical variables, together and by study arm. We will produce scatterplots and histograms in exploratory data analysis, to assess for outliers. As weight heaping is known to be an issue in reporting birthweight and may underreport children born with low birthweight, we will graph distribution of birthweight, and report as a potential limitation of the study.

Effect of treatment group on Bayley and HOME scores: We will use an intention to treat model as our primary outcome, but will also assess per-protocol model as a secondary analysis.

We will test the hypotheses of effect of being randomized to intervention arm on Bayley and HOME score using a three level mixed effects regression as follows:

$$Y_{ijk} = B_0 + dX_{ijk} + u_i + u_{j(i)} + e_{ijk}$$

Where:

Y_{ijk} is Bayley score (or HOME score) in child k, in CHW j in NGO or Anganwadi Center i

d is effect at cluster level

X_{ijk} = treatment assignment for child

u_i = random intercept for AC/NGO

$u_{j(i)}$ = random intercept for AW/CHW within AC/NGO

e_{ijk} = random error term

We will include assessments of effect modification as follows:

$$Y_{ijk} = B_0 + dX_{ijk} + fBas_i + u_i + u_{j(i)} + e_{ijk}$$

where Bas is a vector of time-invariant baseline covariates, if necessary

As an example, heterogeneity of effect by sex will be estimated as follows:

$$Y_{ijk} = B_0 + dX_{ijk} + B_1S_i + (dX_{ijk} * S_i) + u_i + u_{j(i)} + e_{ijk}$$

or

$$Y_{ijk} = B_0 + dX_{ijk} + B_1 S_i + (dX_{ijk} * S_i) + fBas_i + u_i + u_{j(i)} + e_{ijk}$$